(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
24 October 2002 (24.10.2002)

PCT

(10) International Publication Number
WO 02/084643 A1

(51) International Patent Classification: G10L 13/04

(21) International Application Number: PCT/GB02/01277

(22) International Filing Date: 15 March 2002 (15.03.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
01116524.3 11 April 2001 (11.04.2001) CN

(71) Applicant *(for all designated States except US)*: **INTERNATIONAL BUSINESS MACHINES CORPORATION** [US/US]; New Orchard Road, Armonk, NY 10504 (US).

(71) Applicant *(for MG only)*: **IBM UNITED KINGDOM LIMITED** [GB/GB]; P.O. Box 41, North Harbour, Portsmouth, Hampshire PO6 3AU (GB).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: **TANG, Donald** [US/US]; 49 Fox Den Road, Mt. Kisco, NY 10549 (US). **SHEN, LiQin** [CN/CN]; 5-10-09 XinKangYuan XiaoQu, XiSanQi, Beijing 100096 (CN). **SHI, Qin** [CN/CN]; 2-401, 13 No, JuYuan, ShangDi HaiDian District, Beijing 100085 (CN). **ZHANG, Wei** [CN/CN]; Rm. 442, Bld. 25, Fanghui Yuan, Yihe ShangZhuang, Haidian District, Beijing 100095 (CN).

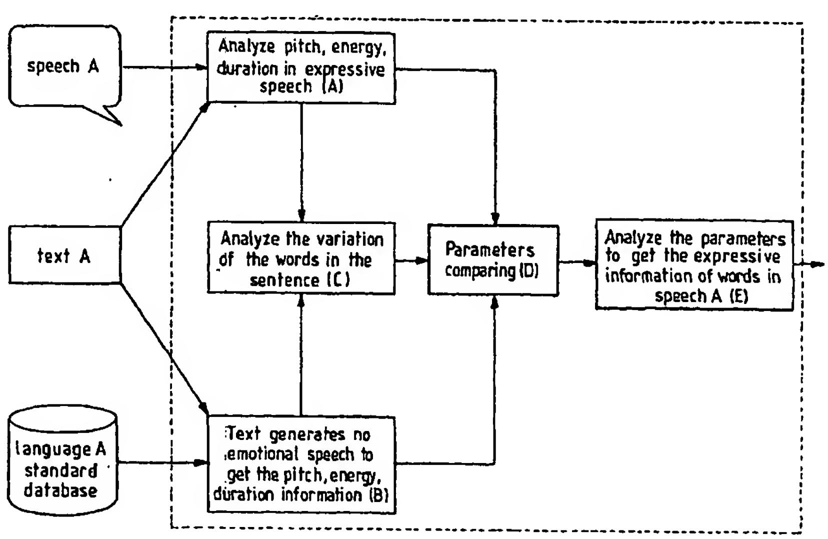
(74) Agent: **WALDNER, Philip**; IBM United Kingdom Limited, Intellectual Property Law, Hursley Park, Winchester, Hampshire SO21 2JN (GB).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZM, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM),

[Continued on next page]

(54) Title: SPEECH-TO-SPEECH GENERATION SYSTEM AND METHOD



(57) Abstract: This invention discloses a expressive speech-to-speech generation system and method. The system and method can generate expressive speech output by using expressive parameters extracted from the original speech signal to drive the standard TTS system. The system comprises: speech recognition means, machine translation means, text-to-speech generation means, expressive parameter detection means for extracting expressive parameters from the speech of language A, and expressive parameter mapping means for mapping the expressive parameters extracted by the expressive parameter detection means from language A to language B, and driving the text-to-speech generation means by the mapping results to synthesize expressive speech. The system and method can improve the quality of the speech output of the translating system or TTS system.

WO 02/084643 A1

European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *with international search report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

SPEECH-TO-SPEECH GENERATION SYSTEM AND METHOD

Field of the Invention

This invention relates generally to the field of machine translation, and in particular to an expressive speech-to-speech generation system and method.

Background of the Invention

Machine translation is a technique to convert the text or speech of a language to that of another language by using a computer. In other words, the machine translation is to automatically translate one language into another language without the involvement of human labor by using the huge memory capacity and digital processing ability of computer to generate dictionary and syntax with mathematics method, based on the theory of language formation and structure analysis.

Generally speaking, current machine translation system is a text-based translation system, which translates the text of one language to that of another language. But with the development of society, the speech-based translation system is needed. By using current speech recognition technique, text-based translation technique and TTS (text-to-speech) technique, a first language speech may be recognized with the speech recognition technique and transformed into the text of the language; then the text of the first language is translated into that of a second language, based on which, the speech of the second language is generated by using the TTS technique.

However, the existing TTS systems usually produce inexpressive and monotonous speech. For a typical TTS system available today, the standard pronunciations of all the words (in syllables) are first recorded and analyzed, and then relevant parameters for standard "expressions" at the word level are stored in a dictionary. A synthesized word is generated from the component syllables, with standard control parameters defined in a dictionary, using the usual smoothing techniques to stitch the components together. Such a speech production cannot create speech that is full of expressions based on the meanings of the sentence and the emotions of the speaker.

Therefore, the embodim nt of th pr sent invention provides an expressive sp ech-to-speech system and method.

According to the embodiment of the present invention, an expressive speech-to-speech system and method uses expressive parameters obtained from the original speech signal to drive a standard TTS system to generate expressive speech.

According to one aspect of the invention there is provided a speech-to-speech generation system as described in claim 1.

According to a second aspect of the invention there is provided a speech-to-speech generation system as described in claim 6.

According to a third aspect of the invention there is provided a method of speech-to-speech generation as described in claim 10.

According to a fourth aspect of the invention there is provided a method of speech-to-speech generation as described in claim 16.

The expressive speech-to-speech system and method of the present embodiment can improve the speech quality of translating system or TTS system.

The aforementioned and further objects and features of the invention could be better illustrated in the following detailed description with accompanying drawings. The detailed description and embodiments are only intended to illustrate the invention.

BRIEF DESCRIPTION OF THE DRAWINGS

Fig. 1 is a block diagram of an expressive speech-to-speech system according to the present invention;

Fig. 2 is a block diagram of an expressive parameter detection means in Fig. 1 according to an embodiment of the present invention;

Fig. 3 is a block diagram showing an expressive parameter mapping means in Fig. 1 according to an embodiment of the present invention;

Fig. 4 is a block diagram showing an expressive speech-to-speech system according to another embodiment of the present invention;

Fig. 5 is a flowchart showing procedures of expressive speech-to-speech translation according to an embodiment of the present invention;

Fig. 6 is a flowchart showing procedures of detecting expressive parameters according to an embodiment of the present invention;

Fig. 7 is a flowchart showing procedures of mapping detecting expressive parameters and adjusting TTS parameters according to an embodiment of the present invention; and

Fig. 8 is a flowchart showing procedures of expressive speech-to-speech translation according to another embodiment of the present invention.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS

As shown in Fig. 1, an expressive speech-to-speech system according to an embodiment of the present invention comprises: speech recognition means 101, machine translation means 102, text-to-speech generation means 103, expressive parameter detection means 104 and expressive parameter mapping means 105. The speech recognition means 101 is used to recognize the speech of language A and create the corresponding text of language A; the machine translation means 102 is used to translate the text from language A to language B; the text-to-speech generation means 103 is used to generate the speech of language B according to the text of language B; the expressive parameter detection means 104 is used to extract expressive parameters from the speech of language A; and the expressive parameters mapping means 105 is used to mapping the expressive parameters extracted by the expressive parameter detection means from language A to language B and drive the text-to-speech generation means by the mapping results to synthesize expressive speech.

As known to those skilled in the art, there are many prior arts to accomplish the Speech Recognition Means, Machine Translation Means and TTS Means. So we only describe expressive parameter detection means and expressive parameter mapping means according to an embodiment of this invention with Fig. 2 and Fig. 3.

Firstly, the key parameters that reflect the expression of speech were introduced.

The key parameters of speech, which control expression, can be defined at different levels.

1. At word level, the key expression parameters are: speed (duration), volume (energy level) and pitch (including range and tone). Since a word generally consists of several characters/syllables (most words have two or more characters/syllables in Chinese), such expression parameters must also be defined at the syllable level, in the form of vectors or timed sequences. For example, when a person speaks angrily, the word volume is very high, the words pitch is higher than normal condition and its envelope is not smooth, and many of pitch mark points even disappear. And at the same time the duration becomes shorter. Another example is that when we speak a sentence in a normal way, we would probably emphasize some words in the sentence, changing the pitch, energy and duration of these words.

2. At sentence level, we focus on the intonation. For example, the envelope of an exclamatory sentence is different from that of a declarative statement.

The following is to describe how the expressive parameter detection means and the expressive parameter mapping means work according to this invention with Fig. 2 and Fig. 3. That is how to extract expressive parameters and use the extracted expressive parameters to drive the text-to-speech generation means to synthesize expressive speech.

As shown in Fig. 2, the expressive parameter detection means of the invention includes the following components:

Part A: Analyze the pitch, duration and volume of the speaker. In Part A, we exploit the result of Speech Recognition to get the alignment result between speech and words (or characters). And record it in the following structure:

```
Sentence Content
{
    Word Number;
    Word Content
{  Text;
    Soundslike;
     Word position;
```

```
 Word property;
Speech start time;
Speech end  time;
*Speech wave;
Speech parameters Content
    { * absolute parameters;
        *relative parameters;
     }
}
}
```

Then we use Short Time Analysis method to get such parameters:

1. Short time energy of each Short Time Window.
2. Detect the pitch contour of the word.
3. The duration of the words.

According these parameters, we take a step forward to get the following parameters:

1. Average Short time energy in the word.
2. Top N short time energy in the word.
3. Pitch range, maximum pitch, minimum pitch, and the value of the pitch in the word.
4. The duration of the word.

Part B: according to the text of the result of speech recognition, we use a standard language A TTS System to generate the speech of language A without expression, and then analyze the parameters of the no expressive TTS. The parameters are the reference of analysis of expressive speech.
Part C: we analyze the variation of the parameters for these words in a sentence forming expressive and standard speech. The reason is that different people speak with different volume and pitch at different speeds. Even for a person, when he speaks the same sentences at different time, these parameters are not the same. So in order to analyze the role of the words in a sentence according to the reference speech, we use the relative parameters.

We use normalized parameter method to get the relative parameters from absolute parameters. The relative parameters are:

1. The relative average Short time energy in the word.

2. Th relative Top N short time energy in the word.

3. The relative Pitch range, relative maximum pitch, relative minimum pitch in the word.

4. The relative duration of the word.

Part D: analyze the expressive speech parameters at word level and at sentence level according to the reference that comes from the standard speech parameters.

1. At the word level, we compare the relative parameters of the expressive speech with those of the reference speech to see which parameters of words vary violently.

2. At the sentence level, we sort the words according to their variation level and word property, and get the key expressive words in the sentences.

Part E: according to the result of parameters comparison and the knowledge that what certain expression will cause what parameters vary, we get the expressive information of the sentence, i.e. detect the expressive parameters, and record the parameter according to the following structure:

```
Expressive information
{
Sentence expressive type;
Words content
  {        Text;
    Expressive type;
    Expressive level;
    *Expressive parameters;
 };
     }
```

For example, when we speak "i•!" angrily in Chinese, many pitches disappear, and the absolute volume is higher than reference and at the same time the relative volume is very sharp, and the duration is much shorter than the reference. Thus, it can be concluded that the expression at the sentence level is angry. The key expressive word is "iš{".

The following is to describe how the expressive parameter mapping means according to an embodiment of this invention is structured, with reference to Fig. 3A and Fig. 3B. The expressive parameter mapping means comprises:

Part A: Mapping the structure of expressive parameters from language A to language B according to the machine translation result. The key method is to find out what words in language B to which the words in language A, which are important for showing expression, correspond. The following is the mapping result:

```
Sentence content for language B
{
       Sentence Expressive type;
word content of  language B
{  Text;
     Soundslike;
     Position in sentence;
     Word expressive information in language A;
     Word expressive information in language B;
}
}

Word expressive of language A
  {          Text;
     Expressive type;
     Expressive level;
     *Expressive parameters;
 }

           Word expressive of language B
  {
Expressive type;
Expressive level;

*Expressive parameters;
}
```

Part B: Based on the mapping result of expressive information, the adjustment parameters that can drive the TTS for language are generated. By this means, we use an expressive parameter table of language B to give

out which words use what a set of parameters according to the expressive parameters. The parameters in the table are the relative adjusting parameters.

The process is shown in Fig. 3B. The expressive parameters are converted by converting tables of two levels (words level converting table and sentence level converting table), and become the parameters for adjusting the text-to-speech generation means.

The converting tables of the two levels are:

1. The word level converting table, for converting expressive parameters to the parameters that adjust TTS.

The following is the structure of the table:

```
Structure of Word TTS adjusting Parameters table
{
      Expressive_Type ;
      Expressive_Para;
         TTS adjusting parameters;
};

Structure of TTS adjusting parameters
{
     float Fsen_P_rate;
      float Fsen_am_rate;
      float Fph_t_rate;
      struct Equation Expressive_equat; ( for changing the curve
characteristic of  pitch contour)
};
```

2. The sentence level converting table, for giving out the prosody parameters of the sentence level according to emotional type of the sentence to adjust the parameters at the word level adjustment TTS.

```
Structure of sentence TTS adjusting Parameters table
{
      Emotion_Type ;
           Words_Position;
           Words_property;
```

```
TTS adjusting parameters;

 };


Structure of TTS adjusting parameters
{
        float Fsen_P_rate;
        float Fsen_am_rate;
        float Fph_t_rate;
        struct Equation Expressive_equat; ( for changing the curve
   characteristic of  pitch contour)
      };
```

The speech-to-speech system according to the present invention has been described as above in connection with embodiments. As known to those skilled in the art, the present invention can also be used to translate different dialects of the same language. As shown in Fig. 4, the system is similar to that in Fig. 1. The only difference is that the translation between different dialects of the same language does not need the machine translation means. In particular, the speech recognition means 101 is used to recognize the speech of language A and create the corresponding text of language A; the text-to-speech generation means 103 is used to generate the speech of language B according to the text of language B; the expressive parameter detection means 104 is used to extract expressive parameters from the speech of dialect A; and the expressive parameter mapping means

105 is used to map the expressive parameters extracted by expressive parameter detection means 104 from dialect A to dialect B and drive the text-to-speech generation means with the mapping results to synthesize expressive speech.

The expressive speech-to-speech system according to the present invention has been described in connection with Fig. 1-4. The system generates expressive speech output by using expressive parameters extracted from the original speech signals to drive the standard TTS system.

The present invention also provides an expressive speech-to-speech method. The following is to describe an embodiment of speech-to-speech translation process according to the invention, with Fig. 5-8.

As shown in Fig. 5, an expressive speech-to-speech method according to an embodiment of the invention comprises the steps of: recognizing the speech of language A and creating the corresponding text of language A (501); translating the text from language A to language B (502); generating the speech of language B according to the text of language B (503); extracting expressive parameters from the speech of language A (504); and mapping the expressive parameters extracted by the detecting steps from language A to language B, and driving the text-to-speech generation process by the mapping results to synthesize expressive speech (505).

The following is to describe the expressive detection process and the expressive mapping process according to an embodiment of the present invention, with Fig. 6 and Fig. 7. That is how to extract expressive parameters and use the extracted expressive parameters to drive the existing TTS process to synthesize expressive speech.
As shown in Fig. 6, the expressive detection process comprises the steps of:

Step 601: analyze the pitch, duration and volume of the speaker. In Step 601, we exploit the result of speech recognition to get the alignment result between speech and words (or characters). Then we use Short Time Analyze method to get such parameters:

1. Short time energy of each Short Time Window.
2. Detect the pitch contour of the word.
3. The duration of the words.

According these parameters, we take a step forward to get the following parameters:

1. Average Short time energy in the word.
2. Top N short time energy in the word.
3. Pitch range, maximum pitch, minimum pitch, and pitch number in the word.
4. The duration of the word.

Step 602: according to the text that is th result of speech recognition, we use a standard language A TTS System to generate the speech of language A without expression. Then analyze the parameters of the inexpressive TTS. The parameters are the reference of analysis of expressive speech.

Step 603: Analyze the variation of the parameters for these words in the sentence that is from expressive and standard speech. The reason is that different people maybe speak with different volume, different pitch, at different speed. Even for a person, when he speaks the same sentences at different time, these parameters are not the same. So in order to analyze the role of the words in the sentence according to the reference speech, we use the relative parameters.

We use normalized parameter method to get the relative parameters from absolute parameters. The relative parameters are:

1. The relative average short time energy in the word.
2. The relative top N short time energy in the word.
3. The relative pitch range, relative maximum pitch, relative minimum pitch in the word.
4. The relative duration of the word.

Step 604: analyze the expressive speech parameters at word level and at sentence level according to the reference that comes from the standard speech parameters.

1. At the word level, we compare the relative parameters of the expressive speech with those of the reference speech to see which parameters of which words vary violently.

2. At the sentence level, we sort the words according to their variation level and word property, to get the key expressive words in the sentences.

Step 605: according to the result of parameters comparison and the knowledge that what certain expression will cause what parameters vary, we get the expressive information of the sentence or in another word, detect the expressive parameters.

Next, we describe the expressive mapping process according to an embodiment of the present invention in connection with Fig. 7. The process comprises steps of:

Step 701: mapping the structure of expressive parameters from language A to language B according to the machine translation result. The key method is to find out the words in language B corresponding to those in language A that are important for expression transfer.

Step 702: according to the mapping result of expressive information, generate the adjusting parameters that could drive language B TTS. By this means, we use an expressive parameter table of language B, according to which the word or syllable synthesis parameters are provided.

The speech-to-speech method according to the present invention has been described in connection with embodiments. As known to those skilled in the art, the present invention can also be used to translate different dialects of the same language. As shown in Fig. 8, the processes are similar to those in Fig. 5. The only difference is that the translation between different dialects of the same language does not need the text translation process. In particular, the process comprises the steps of: recognizing the speech of dialect A, and creating the corresponding text (801); generating the speech of language B according to the text of language B (802); extracting expressive parameters from the speech of dialect A (803); and mapping the expressive parameters extracted by the detecting steps from dialect A to dialect B and then applying the mapping results to the text-to-speech generation process to synthesize expressive speech (804).

The expressive speech-to-speech system and method according to the preferred embodiment have been described in connection with figures. Those having ordinary skill in the art may devise alternative embodiments without departing from the sprit and scope of the present invention. The present invention includes all those modified and alternative embodiments. The scope of the present invention shall be limited by the companying claims.

CLAIMS

1. A speech-to-speech generation system, comprising:

speech recognition means, for recognizing the speech of language A and creating the corresponding text of language A;

machine translation means for translating the text from language A to language B;

text-to-speech generation means, for generating the speech of language B according to the text of language B,

said speech-to-speech translation system is characterized by further comprising:

expressive parameter detection means, for extracting expressive parameters from the speech of language A; and

expressive parameter mapping means for mapping the expressive parameters extracted by the expressive parameter detection means from language A to language B, and driving the text-to-speech generation means by the mapping results to synthesize expressive speech.

2. A system according to claim 1, characterized in that: said expressive parameter detection means extracts the expressive parameters at different levels.

3. A system according to claim 2, characterized in that said expressive parameter detection means extracts the expressive parameters at the word level.

4. A system according to claim 2, characterized in that said expressive parameter detection means extracts the expressive parameters at the sentence level.

5. A system according to any one of claims 1 to 4, characterized in that said expressive parameter mapping means maps the expressive parameters from language A to language B, then converts the expressive parameters of language B into the parameters for adjusting the text-to-speech generation means by the word level converting and the sentence level converting.

6. A speech-to-speech generation system, comprising:

speech recognition means for recognizing the speech of dialect A and creating the corresponding text;

text-to-speech generation means for generating the speech of another dialect B according to the text,

said speech-to-speech generation system is characterized by further comprising:

expressive parameter detection means, for extracting expressive parameters from the speech of dialect A; and

expressive parameter mapping means, for mapping the expressive parameters extracted by the expressive parameter detection means from dialect A to dialect B, and driving the text-to-speech generation means by the mapping results to synthesize expressive speech.

7. A system according to claim 6, characterized in that said expressive parameter detection means extracts the expressive parameters at different levels.

8. A system according to claim 7, characterized in that said expressive parameter detection means extracts the expressive parameters at the word level.

9. A system according to claim 7, characterized in that said expressive parameter detection means extracts the expressive parameters at the sentence level.

10. A system according to any one of claims 6 to 9, characterized in that said expressive mapping means maps the expressive parameters from dialect A to dialect B, then converting the expressive parameters of dialect B into the parameters for adjusting the text-to-speech generation means by the word level converting and the sentence level converting.

11. A speech-to-speech generation method, comprising the steps of: recognizing the speech of language A and creating the corresponding text of language A;

translating the text from language A to language B;

gen rating the speech of language B according to the text of language B,

said expressive speech-to-speech method is characterized by further comprising the steps of:

extracting expressive parameters from the speech of language A; and

mapping the expressive parameters extracted by the detecting steps from language A to language B, and driving the text-to-speech generation process by the mapping results to synthesize expressive speech.

12. A method according to claim 11, characterized in that extracting the expressive parameters is performed at different levels.

13. A method according to claim 12, characterized in that said different levels include the word level.

14. A method according to claim 12, characterized in that said different levels include the sentence level.

15. A method according to any one of claims 11 to 14, characterized in that mapping the expressive parameters from language A to language B, further comprises the step of converting the expressive parameters of language B into the parameters for adjusting the text-to-speech generation means by the word level converting and the sentence level converting.

16. A speech-to-speech generation method, comprising the steps of:

recognizing the speech of dialect A and creating the corresponding text;

generating the speech of another dialect B according to the text, said speech-to-speech generation method is characterized by further comprising steps:

extracting expressive parameters from the speech of dialect A; and

mapping the expressive parameters extracted by the detecting steps from dialect A to dialect B, and driving the text-to-speech generating process by the mapping results to synthesis expressive speech.

17. A method according to claim 16, characterized in that extracting the expressive parameters is performed at different levels.

18. A method according to claim 17, characterized in that said different levels include the word level.

19. A method according to claim 17, characterized in that said different levels include the sentence level.

20. A method according to any one of claims 16 to 19, characterized in that mapping the expressive parameters from dialect A to dialect B, further comprises the step of converting the expressive parameters of dialect B into the parameters for adjusting the text-to-speech generation means by the word level converting and the sentence level converting.

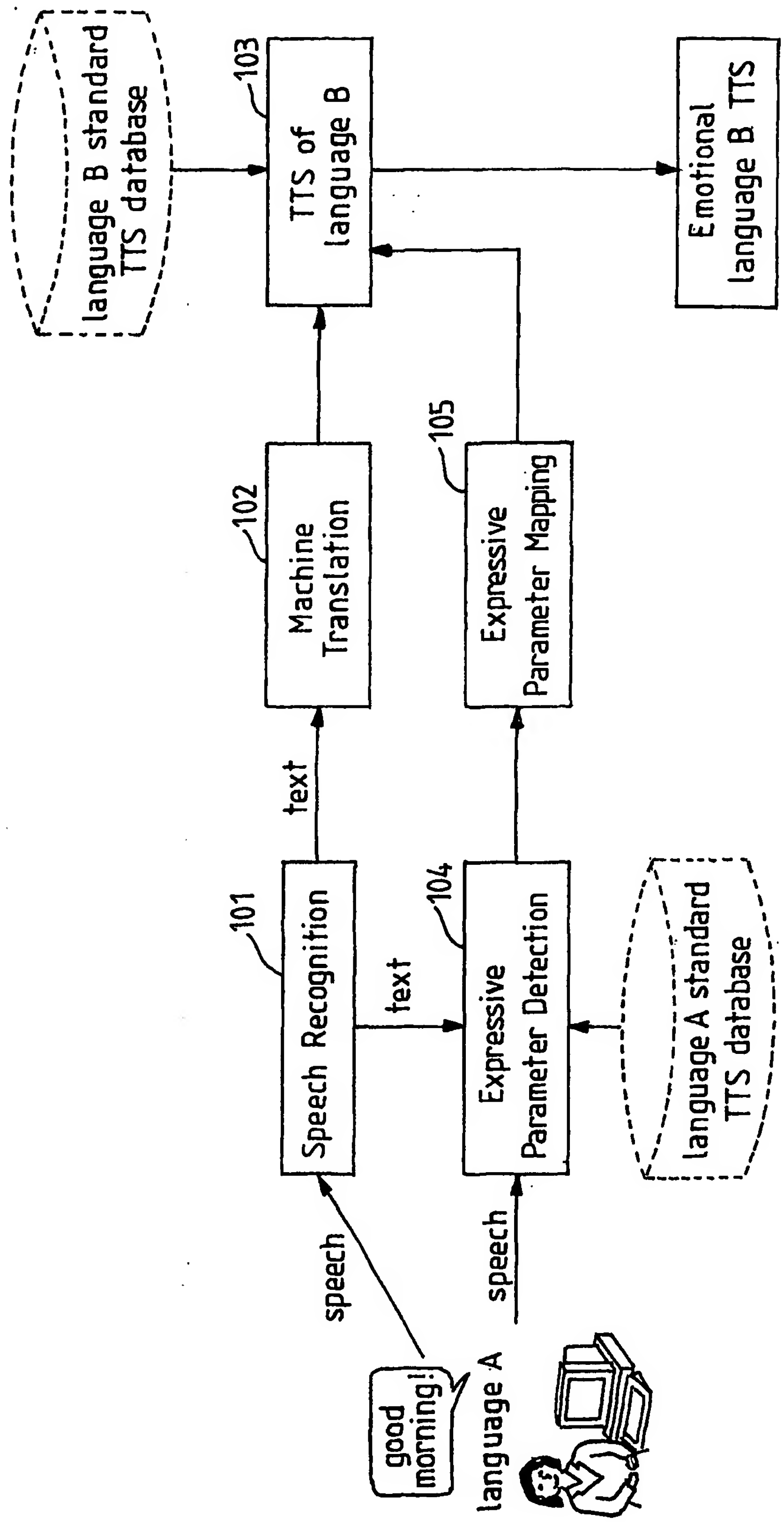

language B standard TTS database
103
TTS of language B
Emotional language B TTS
102
Machine Translation
105
Expressive Parameter Mapping
text
101
Speech Recognition
text
104
Expressive Parameter Detection
language A standard TTS database
speech
speech
good morning!
language A

FIG. 1

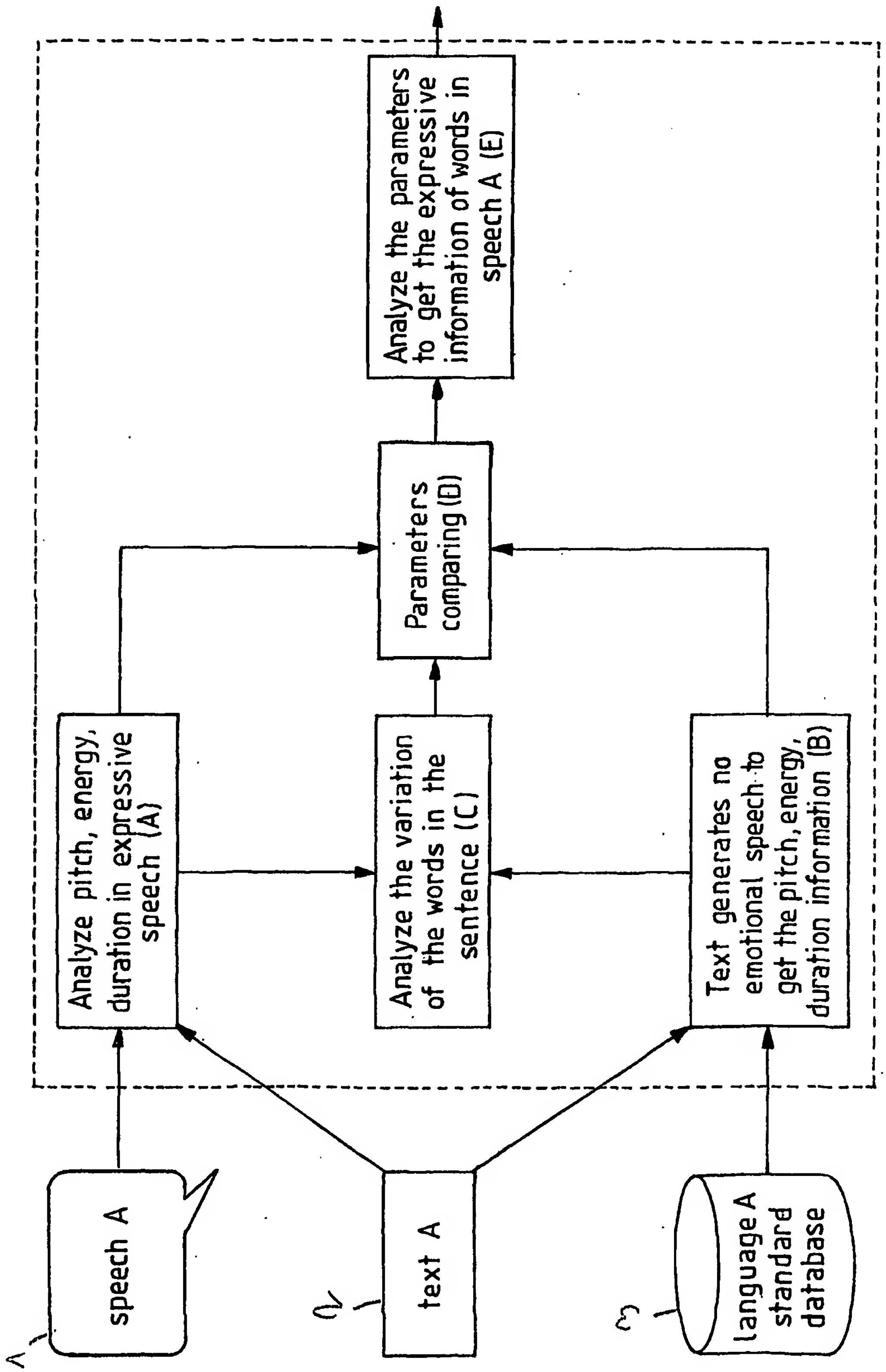

speech A
text A
language A standard database
Analyze pitch, energy, duration in expressive speech (A)
Analyze the variation of the words in the sentence (C)
Text generates no emotional speech to get the pitch, energy, duration information (B)
Parameters comparing (D)
Analyze the parameters to get the expressive information of words in speech A (E)

FIG. 2

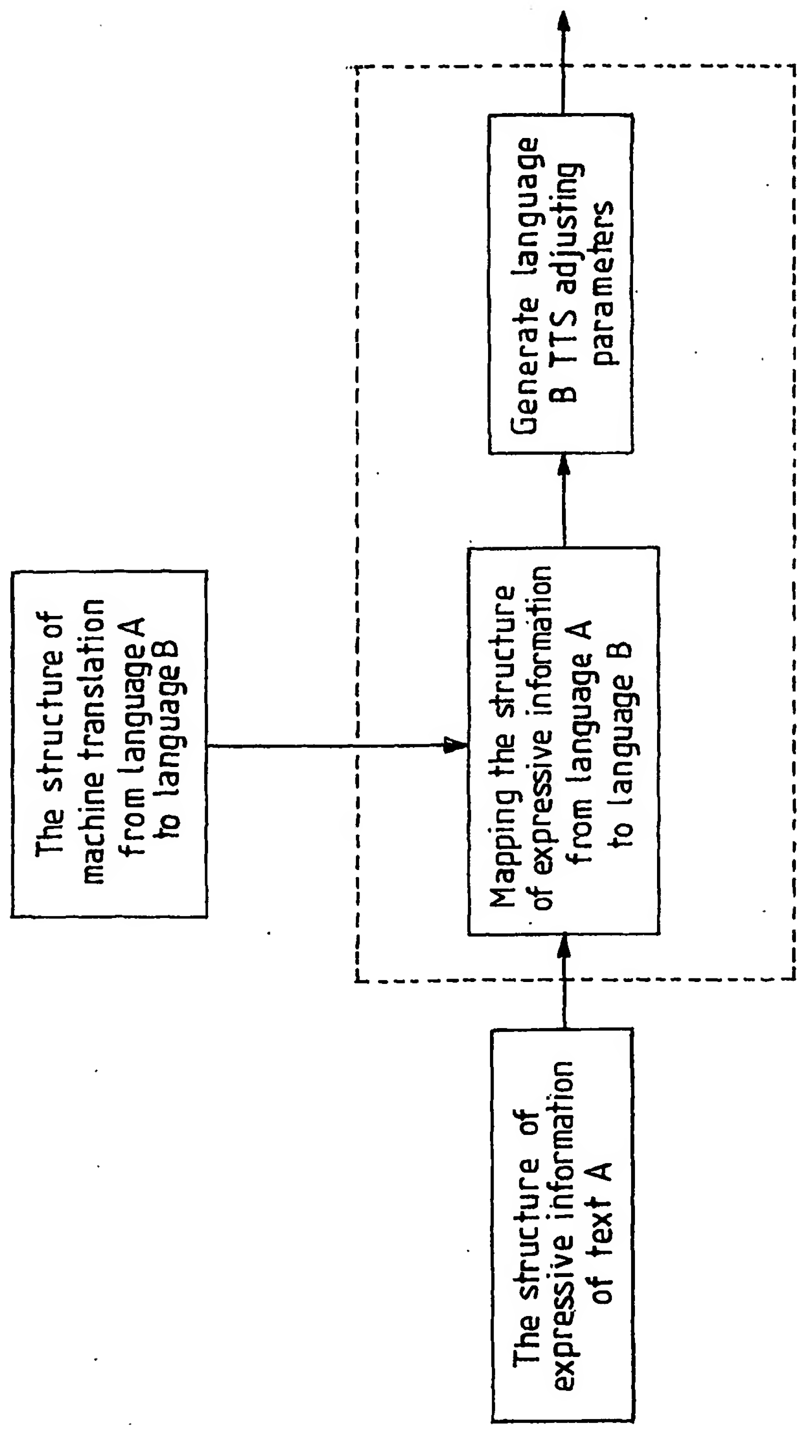

The structure of expressive information of text A
The structure of machine translation from language A to language B
Mapping the structure of expressive information from language A to language B
Generate language B TTS adjusting parameters

FIG. 3A

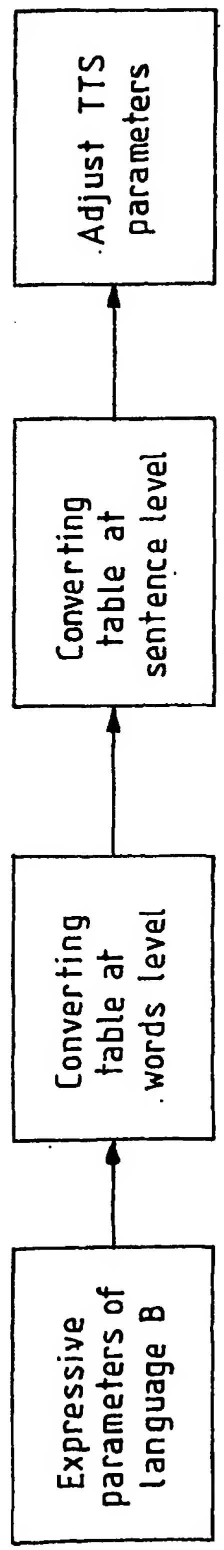

Expressive parameters of language B
Converting table at words level
Converting table at sentence level
Adjust TTS parameters

FIG. 3B

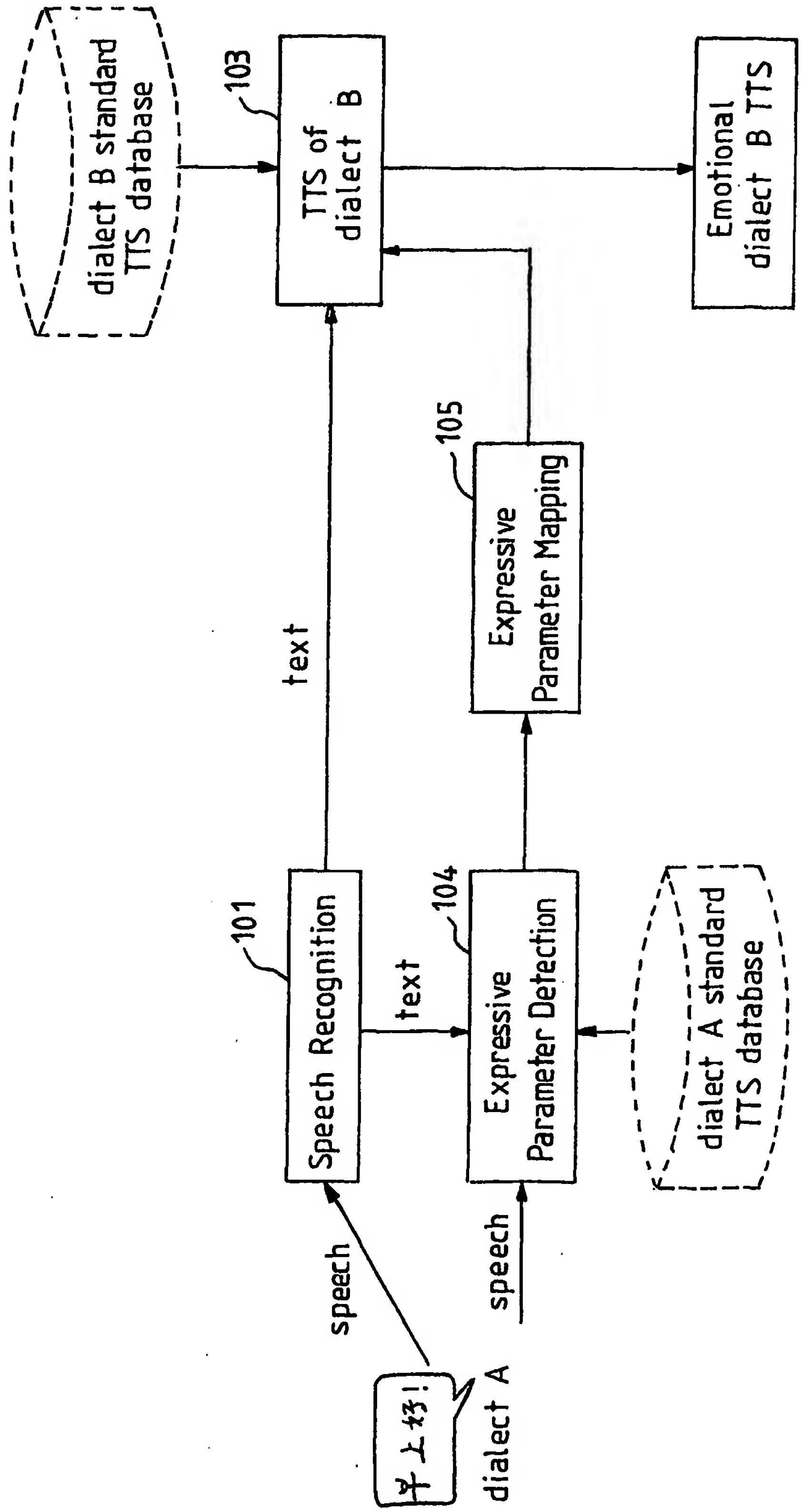

dialect B standard TTS database
103
TTS of dialect B
Emotional dialect B TTS
105
Expressive Parameter Mapping
text
101
Speech Recognition
text
104
Expressive Parameter Detection
dialect A standard TTS database
speech
speech
早上好!
dialect A

FIG. 4

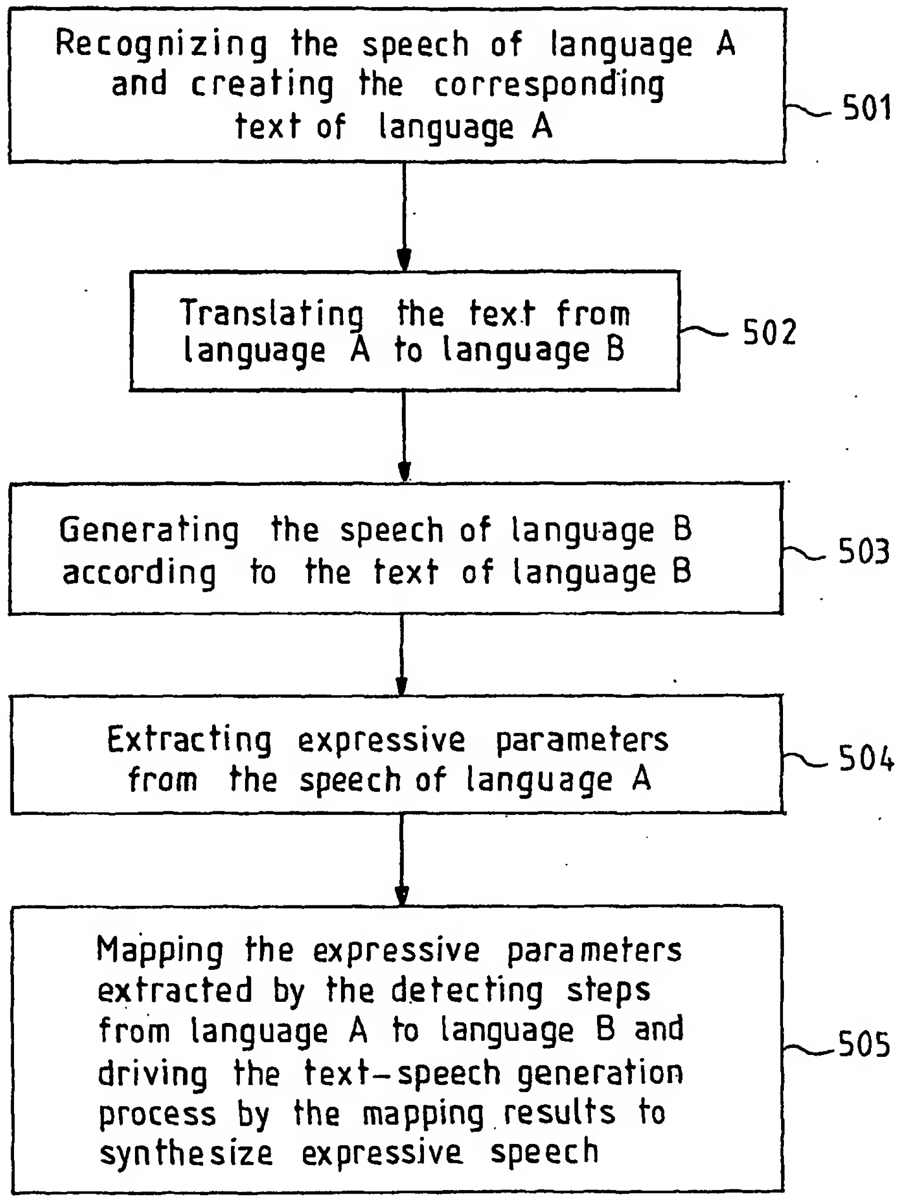

Recognizing the speech of language A and creating the corresponding text of language A
501
Translating the text from language A to language B
502
Generating the speech of language B according to the text of language B
503
Extracting expressive parameters from the speech of language A
504
Mapping the expressive parameters extracted by the detecting steps from language A to language B and driving the text-speech generation process by the mapping results to synthesize expressive speech
505

FIG. 5

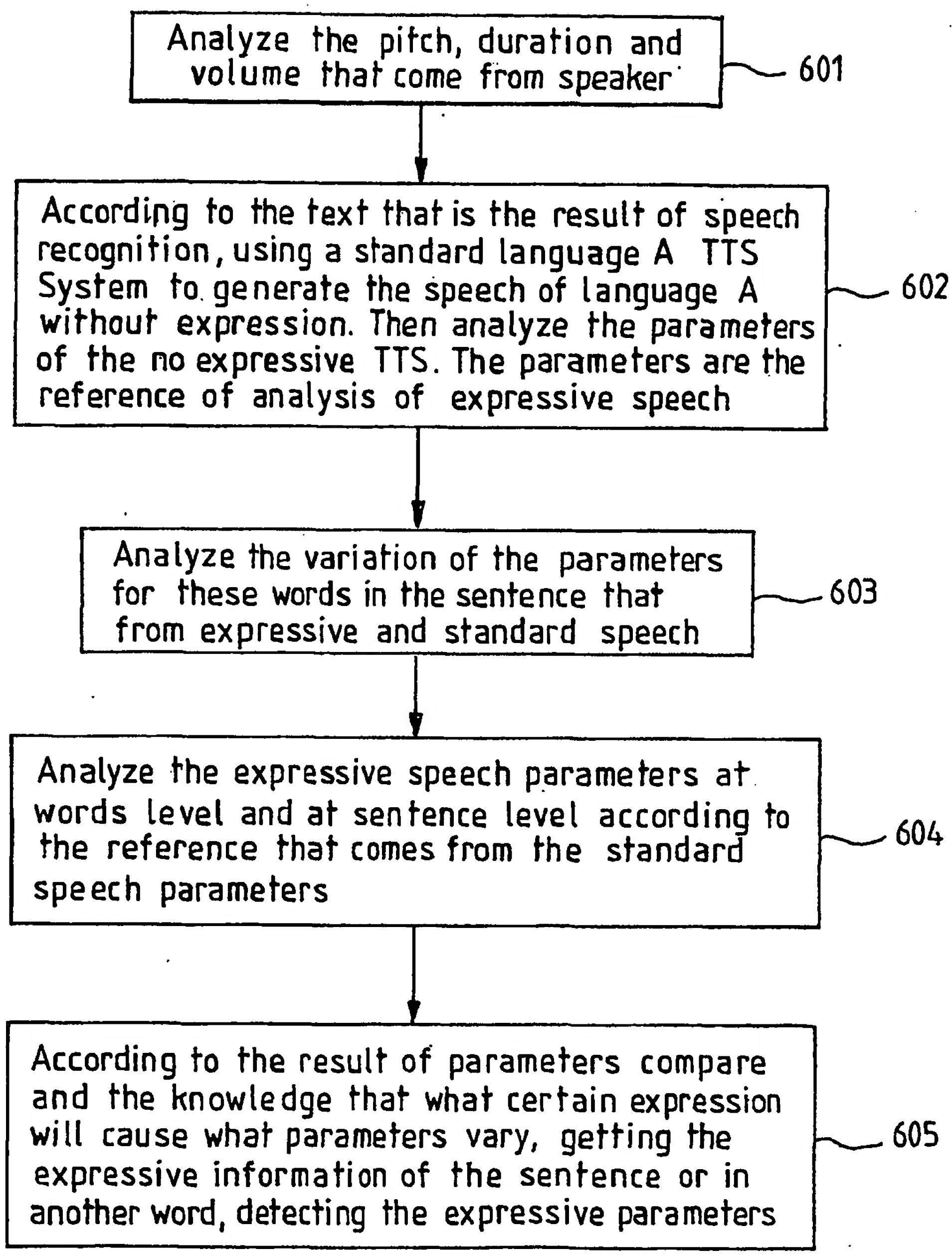

Analyze the pitch, duration and volume that come from speaker
601
According to the text that is the result of speech recognition, using a standard language A TTS System to generate the speech of language A without expression. Then analyze the parameters of the no expressive TTS. The parameters are the reference of analysis of expressive speech
602
Analyze the variation of the parameters for these words in the sentence that from expressive and standard speech
603
Analyze the expressive speech parameters at words level and at sentence level according to the reference that comes from the standard speech parameters
604
According to the result of parameters compare and the knowledge that what certain expression will cause what parameters vary, getting the expressive information of the sentence or in another word, detecting the expressive parameters
605

FIG. 6

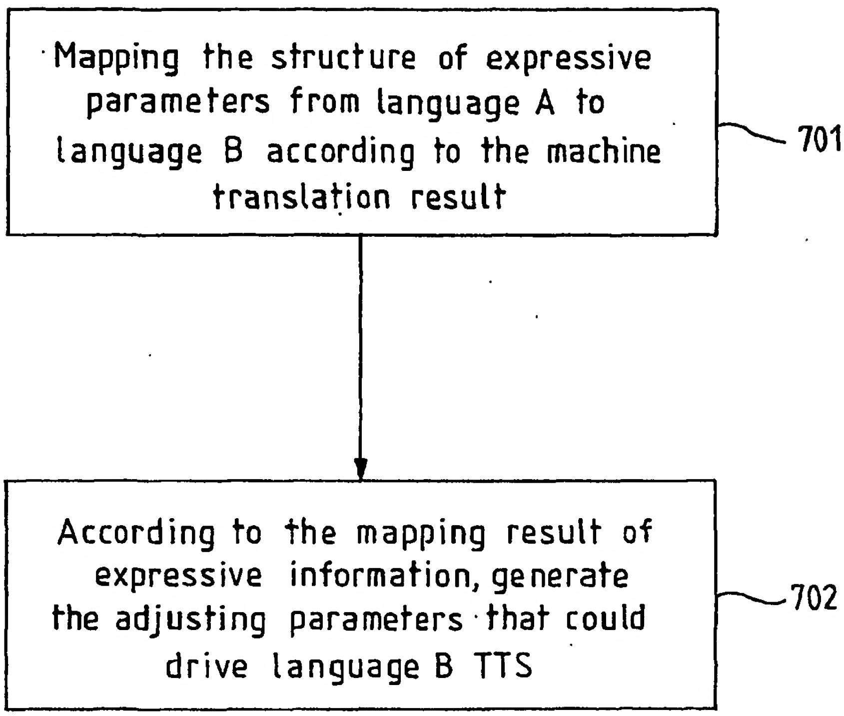

Mapping the structure of expressive parameters from language A to language B according to the machine translation result
701
According to the mapping result of expressive information, generate the adjusting parameters that could drive language B TTS
702

FIG. 7

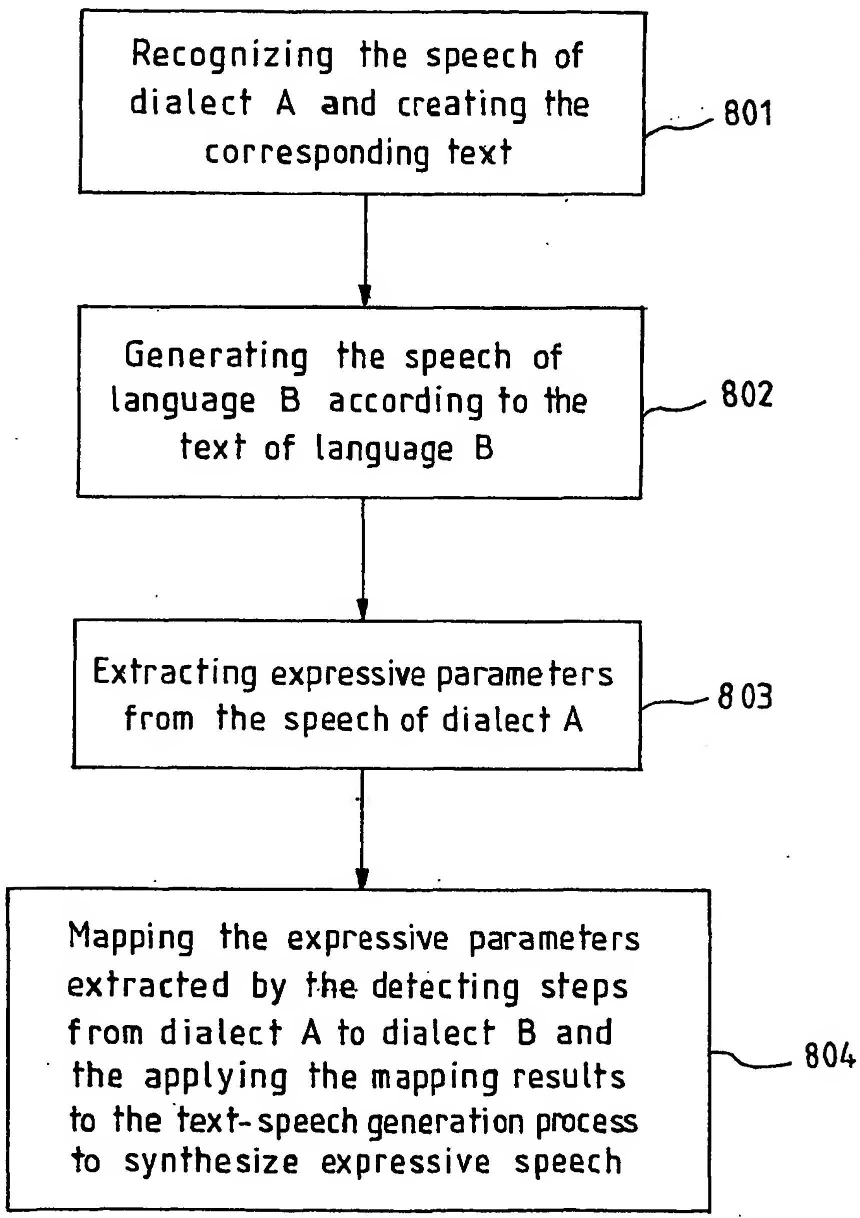

Recognizing the speech of dialect A and creating the corresponding text
801
Generating the speech of language B according to the text of language B
802
Extracting expressive parameters from the speech of dialect A
803
Mapping the expressive parameters extracted by the detecting steps from dialect A to dialect B and the applying the mapping results to the text-speech generation process to synthesize expressive speech
804

FIG. 8

INTERNATIONAL SEARCH REPORT

International Application No
PCT/GB 02/01277

A. CLASSIFICATION OF SUBJECT MATTER
IPC 7 G10L13/04

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 7 G10L G06F

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

WPI Data, PAJ, INSPEC, EPO-Internal

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 97 34292 A (LYBERG BERTIL ;TELIA AB (SE)) 18 September 1997 (1997-09-18) page 2, line 10 -page 3, line 14 page 4, line 20 -page 5, line 15 page 7, line 9 -page 8, line 19 claims 1-15; figure 1	1-20
X	EP 0 624 865 A (TELIA AB) 17 November 1994 (1994-11-17) the whole document	1,6,11, 16

☐ Further documents are listed in the continuation of box C.

☒ Patent family members are listed in annex.

° Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
14 May 2002	24/05/2002

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL – 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Wanzeele, R

INTERNATIONAL SEARCH REPORT

Int   onal Application No
PCT/GB 02/01277

Patent document cited in search report		Publication date	Patent family member(s)		Publication date
WO 9734292	A	18-09-1997	SE	9600959 A	14-09-1997
			WO	9734292 A1	18-09-1997
EP 0624865	A	17-11-1994	SE	500277 C2	24-05-1994
			DE	69420614 D1	21-10-1999
			DE	69420614 T2	06-07-2000
			EP	0624865 A1	17-11-1994
			JP	6332494 A	02-12-1994
			SE	9301596 A	24-05-1994
			US	5546500 A	13-08-1996